Stop Misusing t-SNE and UMAP for Visual Analytics

Hyeon Jeon, Jeongin Park, Sungbok Shin, and Jinwook Seo

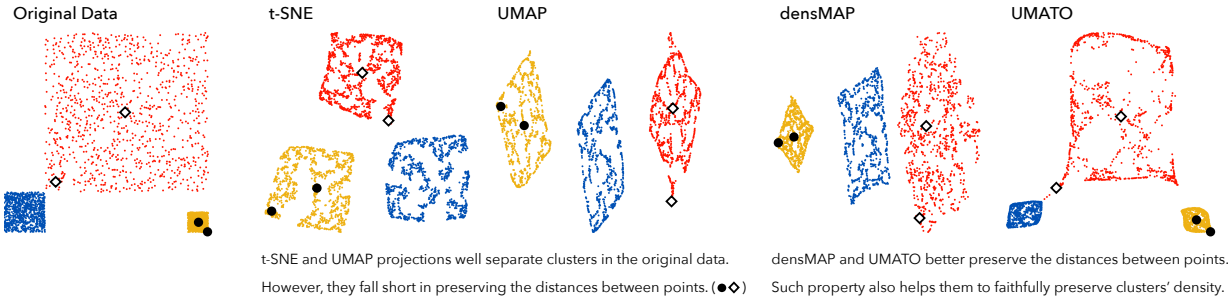

Fig. 1: *Comparison of t-SNE, UMAP, densMAP [54], and UMATO [30] projections of a 2D dataset.* Although *t*-SNE and UMAP do not faithfully represent cluster density or inter-cluster distances, they are often misused to analyze such structures. Our research investigates why this misuse happens and explores strategies to mitigate these issues.

Abstract—Misuses of *t*-SNE and UMAP in visual analytics have become increasingly common. For example, although *t*-SNE and UMAP projections often do not faithfully reflect the original distances between clusters, practitioners frequently use them to investigate inter-cluster relationships. We investigate why this misuse occurs, and discuss methods to prevent it. To that end, we first review 136 papers to verify the prevalence of the misuse. We then interview researchers who have used dimensionality reduction (DR) to understand why such misuse occurs. Finally, we interview DR experts to examine why previous efforts failed to address the misuse. We find that the misuse of *t*-SNE and UMAP stems primarily from limited DR literacy among practitioners, and that existing attempts to address this issue—mostly based on academic papers—have been ineffective. Based on these insights, we discuss potential future research directions to mitigate the misuse.

Index Terms—t-SNE, UMAP, Dimensionality reduction, Misuse, Literature review, Interview study

1 Introduction

We take a close look at a widely known phenomenon threatening the reliability of visual analytics: *the misuse of t-SNE and UMAP*. When practitioners refer to dimensionality reduction (DR) for visually analyzing high-dimensional data, *t*-SNE and UMAP are the most widely applied techniques [10, 17, 33, 48, 76]. However, these two are also commonly misused in practice [7, 8, 13, 73], i.e., *used for analytical tasks that investigate the structure of high-dimensional data that the techniques do not aim to preserve*. For example, although *t*-SNE and UMAP do not accurately represent global structures like distances between points [17, 32, 56, 73], they are often used to investigate the dissimilarity between data points or clusters [7, 8, 73] (Fig. 1). Such misuse may introduce errors into visual analytics, which can propagate through interconnected visualizations and compromise their reliability.

We systematically investigate this misuse to understand how to address it. First, we verify the existence of the misuse by reviewing 136 visual analytics papers that utilize DR. Then, we conduct interviews with 12 researchers who frequently use DR for visual analytics (across data visualization, machine learning, and bioinformatics) to better understand the underlying causes of the misuse. As a final step, we interview eight DR researchers to understand why previous attempts have failed to fully resolve the underlying causes of this misuse.

Our findings indicate that the misuse mainly occurs because practitioners have limited DR literacy. Several participants in our first interview study consider *t*-SNE and UMAP to be "immune to criticism," implying that both authors and reviewers lack sufficient knowledge of how to use DR appropriately. We also find that previous efforts to improve DR literacy have been ineffective. Researchers across various domains produce papers that warn of the weaknesses of *t*-SNE and UMAP and that emphasize how to use DR properly (Sect. 2.2). Unfortunately, such efforts rarely motivate practitioners to meaningfully engage with the issue and hence do not enhance DR literacy.

A plausible direction to mitigate misuse is to leverage automation to guide practitioners with limited DR literacy to use DR properly. However, the optimal level of automation for achieving these goals remains unknown. Fully automating the selection of optimal DR techniques and hyperparameters may help practitioners who are entirely new to DR, but it can discourage them from acquiring DR literacy over time. At the opposite extreme, granting practitioners full agency over DR requires them to configure everything manually, making it both unrealistic and ineffective. Such a tradeoff highlights the need to identify the optimal mixed-initiative approach that leverages machine interventions to help practitioners avoid misuse while preserving their authority.

In summary, our research provides three key contributions:

- We present a **literature review** and **two interview studies** that investigate the misuse of *t*-SNE and UMAP.
- We verify the **prevalence of such misuse** and understand why previous approaches are ineffective in addressing the misuse.
- We discuss **future research directions** to prevent practitioners from misusing DR and to enhance their DR literacy.

We hope this research will spark discussions and encourage the appropriate use of not only DR but also other machine learning techniques, ultimately enhancing the reliability of visual analytics.

- *Hyeon Jeon, Jeongin Park and Jinwook Seo are with Seoul National University. E-mail: {hj, jipark}@hcil.snu.ac.kr, jseo@snu.ac.kr*
- *Sungbok Shin is with Sogang University. E-mail: sbshin90@sogang.ac.kr*
- *Jinwook Seo is the corresponding author.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

2 Backgrounds and Related Work

We discuss how the two DR techniques, *t*-SNE and UMAP, impact visual analytics research. We then detail previous efforts in visualization, machine learning, and bioinformatics fields to address this problem.

2.1 How *t*-SNE and UMAP Impact Visual Analytics

We first explain DR, *t*-SNE, and UMAP. Then, we discuss how the two techniques influence visual analytics.

Dimensionality reduction. DR, e.g., *t*-SNE, UMAP, PCA [57], is an essential tool for visually analyzing high-dimensional data [16, 21, 33, 43, 56]. These techniques receive a high-dimensional dataset as input and produce a 2D representation that preserves important characteristics (e.g., local neighborhood structure or distances between clusters) of the original data. Using DR, any high-dimensional data can be visualized using a single scatterplot. Such effectiveness makes DR widely used for visual analytics across diverse domains [7], including bioinformatics [12], machine learning [25, 39], and finance [11].

***t*-SNE.** Since its first introduction in 2008 [68], *t*-SNE has become one of the most widely used DR techniques. It projects high-dimensional data by minimizing the divergence between two distributions: one representing pairwise similarities of the points in the original space and the other in the low-dimensional space [46, 68].

UMAP. This technique is introduced in 2018 [50] and has quickly gained popularity in diverse fields, including visual analytics. UMAP captures the local structure of high-dimensional data by constructing a *k*-nearest neighbors (*k*NN) graph. It then optimizes a projection by minimizing the cross-entropy between the fuzzy topological representations of *k*NN graphs in the high and low-dimensional spaces.

Impact of *t*-SNE and UMAP in visual analytics. These two techniques significantly influence the visualization and visual analytics fields by motivating numerous follow-up studies. They are frequently utilized in visual analytics systems [10, 48, 76]. These systems typically use projections to provide an overview of the data, allowing users to select subsets via interactions such as brushing and conduct follow-up analysis. However, these techniques primarily preserve local structure and can therefore produce misleading representations of global structure, such as cluster density (Fig. 1). Nevertheless, they are frequently used in ways that risk such misinterpretation. Our work aims to verify this problem by understanding its underlying causes.

2.2 Previous Efforts to Address the Misuse

We identify literature from the visualization, machine learning, and bioinformatics domains that contributes to addressing the misuse of *t*-SNE and UMAP.

Quantitative experiments. These works execute experiments to compare the suitability of the projections generated by diverse DR techniques on different analytic tasks. For example, Xia et al. [75] conduct a user study to test the effectiveness of DR techniques in supporting cluster analysis. They reveal that *t*-SNE and UMAP are most effective for cluster identification tasks but are less effective for tasks such as density or distance comparison. Ventocilla and Reveiro [71] investigate the alignment between human task accuracy in cluster analysis and clustering metrics, reaching a similar conclusion. Jeon et al. [32] and Lause et al. [45] show that *t*-SNE and UMAP work poorly for investigating global structures like cluster density or separability.

Alternative DR techniques. Diverse research proposes new DR techniques as alternatives to *t*-SNE and UMAP that mitigate their weaknesses in representing the original high-dimensional data. Narayan et al. [54] verify that *t*-SNE and UMAP poorly represent cluster density and propose den-SNE and densMAP as alternatives. Trimap [1], PacMAP [72], and UMATO [30] improve UMAP in terms of capturing the global structure (e.g., pairwise distances between data points) of the original data. Global *t*-SNE [78] improves *t*-SNE in the same direction. Fig. 1 illustrates the superiority of alternative techniques against *t*-SNE and UMAP in preserving global structures.

Guidelines for proper use of *t*-SNE and UMAP. These studies inform the limitations of *t*-SNE and UMAP to the public, guiding practitioners to use these techniques more appropriately. Wattenberg et al. [73] provide guidance on appropriately using *t*-SNE, and Coenen and Pearce [13] offer similar insights for UMAP. These works caution practitioners against relying on the global structure presented by *t*-SNE and UMAP and emphasize the substantial impact of hyperparameter selection on the faithfulness of resulting projections. Kobak et al. [42] show that initialization severely affects the faithfulness of the resulting projection and recommend using PCA for initializing *t*-SNE and UMAP. Kobak and Berens [41] recommend using a high learning rate and computing similarity over a diverse range to overcome the limitations of *t*-SNE.

Despite such prior efforts to address misuse, both *t*-SNE and UMAP remain widely misused. We attribute this to practitioners' lack of motivation to cultivate DR literacy. We propose solutions to address this problem, ranging from automating the process of finding optimal projections to actively facilitating discourse on the proper use of DR.

3 Research Objectives

We aim to address the misuse of *t*-SNE and UMAP. We set three research objectives to achieve this goal.

(O1) Verify the misuse of *t*-SNE and UMAP. We want to find evidence that *t*-SNE and UMAP are widely used in visual analytics and are more frequently misused than alternative techniques. We thus want to provide a rationale for focusing on these two techniques.

(O2) Understand why practitioners misuse these techniques. We aim to investigate the underlying cause of the misuse. For example, we want to investigate the rationales behind the misuse of *t*-SNE and UMAP. This investigation grounds our suggestions for future directions to mitigate the misuse.

(O3) Understand why previous efforts fall short in addressing the misuse. We investigate why existing efforts have failed to prevent the persistent misuse of t-SNE and UMAP. As with O2, this investigation supports our suggestion of a new strategy to overcome the limitations of previous approaches.

The remaining parts of this paper are dedicated to achieving these objectives. First, we conduct a literature review (Sect. 4) on visual analytics papers using DR to investigate the extent to which *t*-SNE and UMAP are misused (O1). We deepen our investigation into misuse through an interview study with practitioners who regularly use DR (Sect. 5), observing the underlying causes of the misuse (O2). We then interview expert researchers who study DR (Sect. 6) to obtain insights on why previous approaches are not effective in mitigating the misuse (O3). Based on the findings, we propose future directions to address the misuse of *t*-SNE and UMAP in visual analytics (Sect. 7.1).

4 Literature Review

We execute a literature review to confirm that *t*-SNE and UMAP are commonly misused in visual analytics (O1). We confirm that (**H1**) *t*-SNE and UMAP dominate the use of DR in visual analytics, providing a rationale for our focus on these two techniques. We also hypothesize and verify two types of misuse: (**H2**) the use of *t*-SNE and UMAP for unsuitable tasks, and (**H3**) the lack of appropriate justification.

4.1 Protocol

The review consists of four steps: paper search, categorization, task suitability review, and quantitative analysis.

Paper search. We search for *papers that propose a visual analytics approach, framework, or systems that incorporate the existing DR techniques*. We query papers that satisfy two conditions: (1) the term "`visual analytics`" or "`visual analysis`" appears in the title or abstract, and (2) one or more of the terms among "`dimensionality reduction`", "`dimension reduction`", "`multidimensional projection`", and "`multidimensional scaling`" are mentioned in the full text. We use IEEE Xplore, Wiley online library, and Sagepub to search papers published in major data visualization journals and conferences (e.g., IEEE VIS, TVCG, CG&A, PacificVis, EuroVis, CGF, IVIS). We filter out papers published before 2008, the year *t*-SNE was announced. Then, we inspect each paper and remove those that do

not fall within our search scope (e.g., papers that propose novel DR techniques or pipelines [20, 44] or focus on other kinds of data, e.g., networks [14, 62]). Two authors engage in this process; we exclude a paper only when both agree to minimize false negatives. The eligibility of the papers is determined by the title, abstract, introduction, and methods sections. The list of identified papers is in Appendix A.

Categorization. We categorize the identified papers according to the following three criteria:

DR techniques. We mark a paper as using a certain DR technique only when the paper explicitly mentions that they use the technique to visualize high-dimensional data. We exclude cases in which DR is used solely for data preprocessing or compression [20, 33], as our goal is to determine whether DR techniques, including *t*-SNE and UMAP, are used for unsuitable visual analytics tasks or not (H2).

Target analytic tasks. We categorize a paper as performing a specific analytic task when it states that the task is an intended target of the analysis. Meanwhile, several papers do not explicitly specify their target analytic tasks. In such cases, we examine user studies, case studies, and use cases to infer the target tasks. For instance, if a case study includes analyses that "lasso" a cluster for detailed investigation, we classify the paper as performing a cluster identification.

Rationales. We examine the system design of each paper to identify the stated rationale for selecting particular DR techniques. A paper is marked to have a specific rationale when the rationale is explicitly stated. If no rationale is provided, we label the paper as "No reason".

We select the first two criteria to investigate the extent to which DR techniques are misused for tasks that are not suitable for them (H2). We select rationales as an additional criterion to check whether they justify the use of *t*-SNE, UMAP, and other DR techniques in an inappropriate way (H3). We follow the thematic coding process for the categorization. Two coders first independently categorize the papers, where the initial agreement measured by Cohen's κ is 0.62. They merge and revise their categorizations through three iterative discussions (Sect. 4.2).

Task suitability review. We assess the suitability of major DR techniques for different analytic tasks by examining prior work that evaluates the weaknesses and strengths of different DR techniques [30, 32, 54, 75] (Sect. 2.2). Following our definition of misuse, we consider a DR technique suitable for an analytic task if it preserves the structural characteristics relevant to that task. When suitable DR techniques are used, the corresponding analytic tasks can be conducted more reliably. Two coders conduct the investigation through the following procedure. At first, the coders review the papers independently and record the task suitability of the DR techniques by quoting the texts from the papers. We only consider claims supported by evidence such as computational benchmarks, user studies, or case studies. The coders then unified their findings to derive a final conclusion regarding the task suitability of the DR techniques. If conflicts arise among papers, we prioritize claims supported by stronger evidence (e.g., we place greater weight on findings from user studies than those from case studies) and by a larger number of papers. We depict the results in Sect. 4.3.

Quantitative analysis. We quantitatively analyze the categorized papers to verify H1 and H2. Detailed results are presented in Sect. 4.4.

4.2 Paper Search and Categorization Results

We retrieve 312 papers, and after screening and filtering, we retain 136 papers. The categories we identify from them are described below.

4.2.1 DR Techniques

We identify 18 DR techniques in total. Among them, we find four commonly used techniques (*t*-SNE, UMAP, PCA, and MDS) for which each is used more than 20 times. The other 14 techniques are used fewer than five times each, and we categorize them all under "others."

4.2.2 Analytic Tasks

Our review reveals that analytic tasks using DR can be divided into seven categories (Fig. 2). The first three are *identification* tasks: tasks that find or extract specific data points or their set. The other four are *investigation* tasks: the tasks that measure or quantify characteristics of the data. Detailed descriptions of each task are as follows.

Table 1: *The coverage of tasks (rows) by references (columns).* We refer to each work using the last name of the first author. (Xia: Xia et al. [75], Ete.: Etemadpour et al. [18], Bre.: Brehmer et al. [6], Non.: Nonato and Aupetit [56], Sed.: Sedlmair et al. [61], Cas. : Cashman et al. [7])

Task	Xia	Ete.	Bre.	Non.	Sed.	Cas.
Neighborhood Identification		•		•		•
Outlier Identification	•	•		•		•
Cluster Identification	•	•	•	•	•	•
Point Distance Investigation				•		•
Class Separability Investigation			•	•	•	•
Cluster Distance Investigation	•	•	•	•	•	•
Cluster Density Investigation	•	•		•		

Neighborhood identification. This task aims to find data points similar to a target point based on the proximity within the projection. Since this task supports many other analytic tasks, such as cluster identification, preserving local neighborhood structure is often considered the most important criterion when evaluating DR projections [30, 52, 56].

Outlier identification. This task is about identifying outliers within projections. Analysts often count the number of outliers in the data [18] or determine whether a point is a cluster member or an outlier [75]. The task is typically used to examine the quality of class labels by identifying points with high uncertainty about their membership [26].

Cluster identification. This task involves identifying clusters within DR projections. Analysts typically count the number of clusters [18] or label clusters interactively using selection tools such as lasso or box-shaped brushes [51, 75]. This task often includes investigating subclusters within existing clusters [18]. Visual analytics systems often provide auxiliary visualizations to show detailed information about the identified clusters [9, 48].

Point distance investigation. Similar to the cluster distance investigation task, this task investigates the distance between data points as a proxy for their high-dimensional dissimilarity. It can be interpreted as a "continuous" version of the neighborhood identification task.

Class separability investigation. This task investigates how distinctly different classes are separated or distinct in the projections, where the classes are color-coded. It involves analyzing both the degree to which classes are mixed with each other and the relative distances among classes within the projection. The task is commonly performed when DR techniques are employed to explain the behavior of a supervised machine learning model, particularly to illustrate how the model distinguishes between different classes [59].

Cluster distance investigation. This task uses the distance between well-separated clusters as a proxy for their similarity in the original high-dimensional space. The clusters can be explicitly labeled (i.e., color-coded) or implicitly represented by data distribution [61].

Cluster density investigation. This task identifies and compares the density of clusters using cluster density as a proxy for the variability of data points within each cluster [54].

Task coverage validation. To validate the comprehensiveness of our categorization, we examine whether the analytic tasks in our list are included in prior task taxonomies within the visualization field. We review Etemadpour et al. [18], Xia et al. [75], Brehmer et al. [6], Nonato and Aupetit [56], Sedlmair et al. [61], and Cashman et al. [7]. We find that all tasks are covered by at least two previous studies (Table 1), confirming that our categorization covers all important tasks in DR.

4.2.3 Rationales

We identify seven large categories of rationales used to justify the selection of DR techniques. We define each rationale in Table 2. It is worth noting that a substantial number of papers (44%; Fig. 6) do not mention specific rationale, i.e., marked as "No reason".

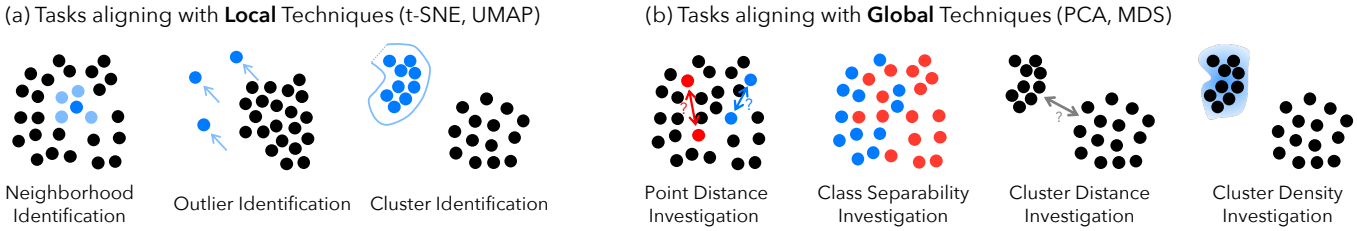

Fig. 2: *Illustrations of the analytic tasks using DR and their alignment to local and global DR techniques.* Our literature review identifies seven types of analytic tasks. *t*-SNE and UMAP are suitable for neighborhood, outlier, and cluster identification tasks, but are inappropriate for other tasks.

Table 2: *The definition of the rationales that we identify from the literature review (Sect. 4.2.3)* Except for extensibility and simplicity, all rationales are leveraged to justify the use of *t*-SNE and UMAP.

Rationale	Definition
Faithfulness	The degree to which DR techniques accurately represent the original structure of the high-dimensional data without distortions.
Popularity	The degree to which DR techniques are widely known and used by practitioners in the visual analytics field.
Scalability	Computational efficiency in executing DR techniques.
Interpretability	The degree to which DR techniques yield visually distinct, analyzable clusters, enabling clear explanation of the data.
Stability	The degree to which DR techniques produce projections that are stable against hyperparameter change or the stochastic nature of DR.
Extensibility	The degree to which DR techniques can be adapted or expanded to accommodate diverse data conditions or input formats.
Simplicity	The degree to which practitioners can readily understand and apply DR techniques.

Faithfulness. Researchers justify the use of DR techniques based on their faithfulness, i.e., their ability to accurately represent the structure of high-dimensional data without distortion. This rationale mostly relies on references to benchmark studies that evaluate DR techniques [17,75]. One notable finding is that researchers often cite the original UMAP paper [50] to support claims about its capability to preserve global structure, which is not always correct [13, 30, 35, 54, 75]. We also find several papers claiming the faithfulness of UMAP without references.

Popularity. Researchers justify the use of DR techniques based on their popularity, which indicates the degree to which the techniques are widely acquainted and used by practitioners in the visual analytics field. For example, papers mention employing *t*-SNE because it is a "default option" in visualizing high-dimensional data or is "widely recognized" by the research community. These papers also highlight specific research communities, such as biology, computer vision, and document clustering, where these techniques are commonly used [7].

Scalability. The use of DR techniques is also justified by their computational efficiency, which enhances the responsiveness of visual analytics systems. For example, papers state that efficient GPU implementations [55, 58] facilitate the effectiveness of DR techniques.

Interpretability. Researchers use DR techniques because they enable a clear explanation of the data with projections that contain visually distinct and, thus, analyzable clusters. This finding aligns with the work by Morariu et al. [53] and Doh et al. [15] that identify clear cluster separation as a key factor influencing the preference for DR projections.

Stability. Researchers justify the selection of DR techniques by highlighting their stability (i.e., the degree to which DR techniques produce projections that are stable against hyperparameter change or the stochastic nature of DR) as a means to improve the reproducibility of their research. For example, one paper argues that *t*-SNE is stable due to its non-convex optimization [2].

Extensibility. We find that researchers justify the use of DR techniques based on their extensibility or their ability to adapt or expand to accommodate diverse data conditions or input formats. For example, some papers use DR techniques because they are parametric, i.e., support new data points to be dynamically projected after initial projection [60], particularly for analyzing streaming, online data.

Simplicity. A few papers mention selecting DR techniques that are simple and easy for practitioners to understand and apply. Among the four major techniques, only PCA has been justified by this rationale.

4.3 Suitability of DR Techniques to Analytic Tasks

We assess the suitability of four major DR techniques (*t*-SNE, UMAP, PCA, and MDS) to the analytic tasks identified in Sect. 4.2.2. This is done by revisiting previous studies that evaluate DR techniques and analyze the alignment between the DR techniques and analytic tasks (Sect. 2.2) (e.g., [17, 30, 52, 56, 75, 78]). This analysis helps examine whether researchers are applying DR to tasks that are suitable (H1).

Here, we clarify that the suitability of DR projections cannot be assessed solely based on the selection of DR techniques. For example, hyperparameters such as perplexity in *t*-SNE can significantly affect how well a projection aligns with different analytic tasks. Still, we adopt a coarse-grained approach that assesses suitability primarily by selecting DR techniques for the following reasons. First, we identify that nearly all studies we investigated did not report their hyperparameter settings or other configurations, like seed selection. Second, the choice of DR technique remains the most influential factor affecting the task suitability of DR projections [17, 56]. Exploring task suitability by considering a broader range of factors would be an interesting avenue for future work. We discuss this direction in Sect. 7.2.

4.3.1 Tasks Suitable for *t*-SNE and UMAP

t-SNE and UMAP focus on preserving local neighborhoods by placing nearby points close together while separating non-neighbors. They are thus commonly referred to as local techniques. Several studies demonstrate that they show state-of-the-art performance in preserving local structures, both empirically [17, 30, 52] and theoretically [46]. We categorize the following tasks that are suitable for *t*-SNE and UMAP.

Neighborhood identification task is more suitable for *t*-SNE and UMAP. As aforementioned, *t*-SNE and UMAP directly aim to preserve local neighborhood structure. This makes them better suited for neighborhood identification tasks than alternative techniques by design [1, 32, 70, 78], which also have been empirically validated [30, 52].

Outlier identification task is more suitable for *t*-SNE and UMAP. Since projections generated by local techniques clearly distinguish neighboring and non-neighboring points, they can effectively separate outliers from clusters. Xia et al. [75] empirically show that *t*-SNE and UMAP are the most effective DR techniques for outlier identification, outperforming alternatives such as PCA.

Cluster identification task is more suitable for *t*-SNE and UMAP. As *t*-SNE and UMAP locate neighboring points close and non-neighboring points far away [30, 52, 70], they clearly represent individual high-dimensional clusters as 2D clusters, thus suitable for the

cluster identification task. Xia et al. [75] show that the participants perform best when identifying clusters with *t*-SNE and UMAP.

4.3.2 Tasks Suitable for PCA and MDS

PCA and MDS are DR techniques that preserve global pairwise distances between data points more effectively than local techniques [30, 56, 67, 69, 75]. They are usually referred to as global techniques. The following tasks are suitable for these DR techniques.

Point distance investigation task is more suitable for PCA and MDS. These techniques are designed to directly preserve the pairwise distance structures more effectively compared to local techniques. They are thus more suitable than *t*-SNE and UMAP in investigating distances between data points (Fig. 1 dots and diamonds). Several studies [1, 30, 52] propose techniques that improve UMAP in preserving global pairwise distances between points, such as UMATO [30] and TriMap [1].

Class separability investigation task is more suitable for PCA and MDS. The superiority of PCA and MDS in preserving distances between data points makes them more precisely exhibit the separability between class labels [32, 73, 74]. Local techniques like *t*-SNE and UMAP can well depict the extent to which classes are "mixed" [32]. However, they are widely reported to exaggerate the distance between classes [3–5, 32]. Wattenberg et al. [73] show that hyperparameter choices can significantly distort class distances in *t*-SNE.

Cluster distance investigation task is more suitable for PCA and MDS. PCA and MDS better preserve pairwise distances between points within each cluster compared to alternatives, making the inter-cluster distances in the resulting projections meaningful. These techniques are thus suitable for tasks involving cluster distance analysis. Many computational benchmarks have validated the superiority of PCA and MDS in supporting the cluster distance investigation task [30, 32, 52, 71, 72]. This implies the appropriateness of these techniques for supporting cluster distance investigation. Xia et al. [75] empirically show that global techniques like PCA enable users to perform this task more accurately than local techniques. In contrast, Wattenberg et al. [73] and Coenen et al. [13] also inform that the distance between clusters lacks meaning in *t*-SNE and UMAP projections, respectively.

Cluster density investigation task is more suitable for PCA and MDS. These techniques depict the similarity between data points as low-dimensional proximity and thus can more sensitively depict the differences in cluster densities. In contrast, local techniques like *t*-SNE and UMAP poorly reflect their true similarity in high-dimensional space [1, 30, 54] as they only focus on neighboring points. As a result, *t*-SNE and UMAP projections poorly represent cluster density (Fig. 1), which motivates the development of improved techniques such as den-SNE and densMAP [54]. The superiority of global techniques in supporting the density investigation task is also validated by Xia et al. [75] through user studies and Jeon et al. [32] via case studies.

4.3.3 Validity of the Suitability Analysis

One notable finding in our suitability analysis is that *t*-SNE and UMAP perform better on all "identification" tasks while PCA and MDS excel at "investigation" tasks. This result aligns with the fundamental differences in how these methods interpret distances. *t*-SNE and UMAP prioritize preserving local neighborhoods by treating similarity as a binary function (neighbors or non-neighbors), making them well-suited to tasks that require identifying distinct clusters or groups. In contrast, PCA and MDS interpret distances as continuous values, enabling more accurate interpretation of relative distances between points. This finding supports the validity of our task categorization in capturing the alignment between DR techniques and the tasks that are suitable for.

4.4 Findings

We present the findings from the analysis of the identified papers (Sect. 4.2). We reveal that *t*-SNE and UMAP are the most commonly adopted DR techniques (H1). Yet, researchers often use them across any tasks, making them simultaneously the most commonly misused DR techniques (H2). We also observe that many papers leverage *t*-SNE and UMAP without justifications or with improper rationale (H3).

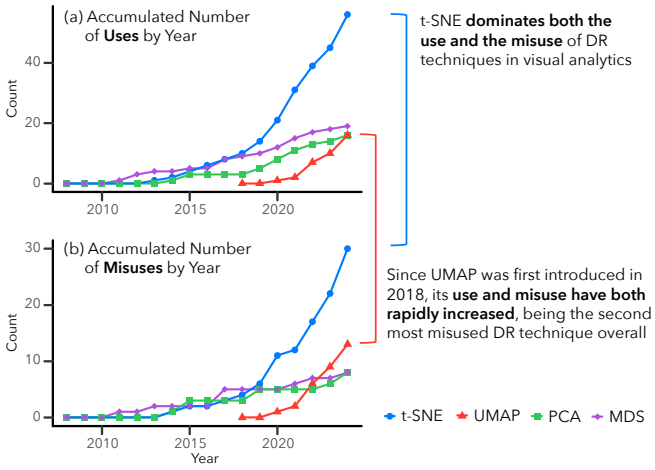

Fig. 3: *The trend of the accumulated number of papers that use (a) or misuse (b) of four major DR techniques.* We collect papers published from 2008, the year *t*-SNE is introduced. UMAP's data also starts from the year it is released (2018). As these data are drawn from the set of papers we surveyed, the actual prevalence of DR use and misuse in practice may be higher than the reported numbers suggest.

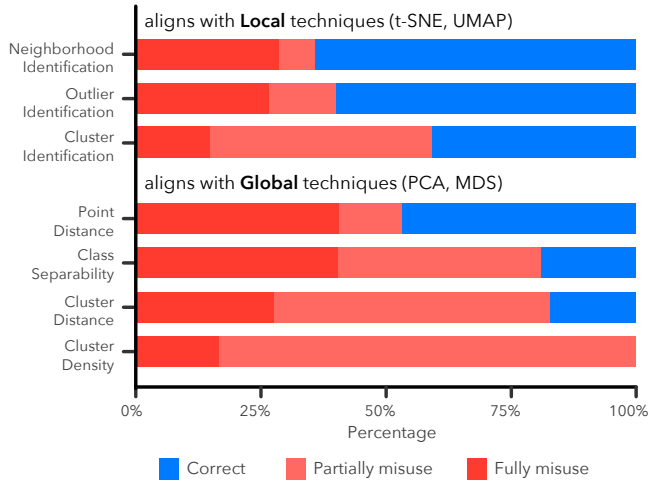

Fig. 4: *The ratio of appropriate use and misuse of DR techniques by analytic tasks.* DR is properly used for tasks that align with local techniques (top 3) but not for those that align with global techniques (bottom 4). This result indicates that *t*-SNE and UMAP are used even for unsuitable tasks.

(Finding 1) *t*-SNE and UMAP dominate the use of DR. While the number of papers using DR has increased over the years, this growth is largely driven by *t*-SNE and UMAP (Fig. 3; H1). *t*-SNE appears in more than half of the identified papers (75 out of 136), more than twice as often as the runner-up. UMAP is used in 31 papers. However, UMAP's adoption is increasing at a much steeper rate than PCA and MDS, enabling it to achieve parity with these established techniques in only six years. These findings highlight that misusing *t*-SNE and UMAP can have a more harmful impact than misusing other techniques.

(Finding 2) *t*-SNE and UMAP are used for *any* tasks. We identify that *t*-SNE and UMAP are commonly used for both suitable and unsuitable tasks, confirming H2. To do so, we determine whether a paper properly uses DR techniques for suitable tasks. Based on our definition of misuse, we mark the paper to be "correct" if all DR techniques used in the paper align with the tasks. We mark the paper to "fully misuse" or "partially misuse" DR when entire or part of DR techniques are used for unsuitable tasks, respectively (See Appendix D for examples). We do so by carefully reviewing the justification for selecting each DR technique, the main task descriptions, and the target tasks described in user and

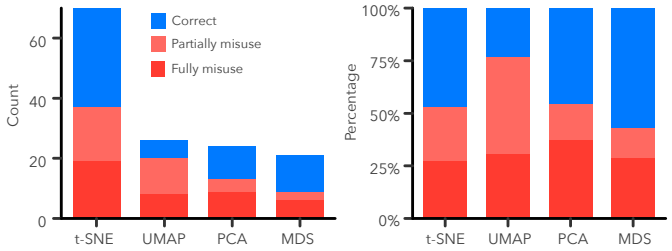

Fig. 5: *The number of appropriate uses and misuses of DR by techniques (left) and their ratio (right).* The analysis reveals that *t*-SNE and UMAP dominate the misuse of DR.

case studies. Then, for each analytic task, we compute the proportion of papers that employ suitable DR techniques relative to all papers addressing the task. As a result (Fig. 4), we find that tasks supported by local techniques have a high proportion of proper usage. However, tasks requiring global techniques have a substantially lower rate of proper usage. This indicates that researchers properly employ local techniques (*t*-SNE and UMAP) when required, while also correctly avoiding global techniques for these tasks. However, tasks requiring global techniques have a substantially lower rate of proper usage, indicating that *t*-SNE and UMAP are misused even for unsuitable tasks.

This tendency causes *t*-SNE and UMAP to dominate the misuse of DR techniques in practice. We find that *t*-SNE is the most widely misused DR technique in our list of papers (Fig. 5a). The misuse of UMAP has recently increased rapidly, making it the runner-up (Fig. 3b). We also find that UMAP has the highest misuse-to-usage ratio among the four major DR techniques (Fig. 5b). In summary, *t*-SNE and UMAP are misapplied more frequently than other techniques, underscoring the need for increased efforts to address these misuses.

In addition, we observe that a similar pattern persists when we exclude papers published before 2016 and 2019—the years when two influential papers guiding researchers on the use of *t*-SNE and UMAP are released [13, 73] (Appendix C). This indicates that the misuse of *t*-SNE and UMAP persists despite existing efforts to inform practitioners how to use these techniques properly.

(Finding 3) *t*-SNE and UMAP are used without rationales or with improper rationales. We find that more than 40% of papers do not explicitly justify their choice of DR techniques (Fig. 6; H3), and this trend persists for *t*-SNE and UMAP. These papers often discuss the general need for DR or describe the techniques' characteristics rather than explaining why specific techniques are chosen. This result implies that practitioners may perceive DR technique selection—including the case of *t*-SNE and UMAP—as requiring less critical evaluation, suggesting a lack of clear understanding of the appropriate way of using DR. Our subsequent interviews (Sect. 5) further reinforce this observation in the context of *t*-SNE and UMAP usage.

We also identify that faithfulness is more widely used to justify the use of *t*-SNE and UMAP compared to PCA and MDS, but is referred to even when these techniques are misused (Fig. 6). This result thus supports our implications that researchers often lack proper understanding of DR; they may not know that *t*-SNE and UMAP are faithful in preserving local structure but not in global structure (Sect. 4.3).

4.5 Takeaways

The following are key takeaways from our literature review.

Misuse of *t*-SNE and UMAP exists. We quantitatively verify that *t*-SNE and UMAP are widely misused in practice (Findings 2 and 3).

Researchers may have a limited understanding of how to use DR properly. We find that researchers often use *t*-SNE and UMAP without proper justification or regard them as faithful for unsuitable tasks (Finding 3), implying limited awareness of various DR techniques and their suitability to analytic tasks. Indeed, the misuse of *t*-SNE and UMAP itself suggests that practitioners do not know how to properly use DR. The common misuse of PCA and MDS, affecting around 50% (Fig. 5), further indicates that this issue is not limited to *t*-SNE and UMAP, but reflects a broader lack of DR literacy. Still, the misuse of *t*-SNE and UMAP is particularly salient because of their popularity. Our interview study (Sect. 5) reaffirms this observation.

Researchers may lack the motivation to use *t*-SNE and UMAP properly. We find that papers misusing *t*-SNE and UMAP have passed peer review and been published in major visualization conferences and journals. This finding suggests that reviewers frequently overlook the importance of using DR appropriately. Furthermore, the frequent absence of clear justifications for selecting DR techniques (Finding 3) implies that researchers often do not recognize their importance.

Table 3: *The demographics of the participants in our interview study with practitioners.* VA and DR denote years of experience in visual analytics and DR, respectively.

	Occupation	Age	Gender	Domain	VA	DR
P1	Professor	35	Male	Visual Analytics	10	7
P2	Undergraduate	22	Female	Visual Analytics	2	1
P3	Research Scientist	30	Male	Computer Vision	·	6
P4	Research Scientist	30	Male	Signal Processing	·	4
P5	Ph.D. Student	28	Female	Visual Analytics	6	5
P6	Ph.D. Student	28	Male	HCI	·	4
P7	Ph.D. Student	29	Male	Chemistry	·	4
P8	Ph.D. Student	24	Female	Visual Analytics	3	2
P9	Research Scientist	30	Male	NLP	·	7
P10	MS Student	22	Male	Visual Analytics	2	1
P11	Staff Engineer	34	Male	Visual Analytics	2	2
P12	Postdoc.	34	Male	Bioinformatics	·	4

5 Interview Study with Practitioners

We aim to investigate why practitioners misuse *t*-SNE and UMAP (O2). To this end, we conduct interviews with practitioners who regularly use DR techniques for their research or visual analytics work.

5.1 Study Design

Participants and recruitment. We want to diversify our participants' experience with DR. We first aim to achieve diversity across the domains in which participants work. To do so, we recruit both visual analytics researchers and domain researchers with experience visually analyzing and presenting their data using *t*-SNE and UMAP. For visual analytics researchers, we randomly select papers from our literature review (Sect. 4), prioritizing diversity in target data and problem domains. We then contact either the first or the corresponding author via email to increase diversity in participants' expertise and visualization literacy. For domain researchers, we ensure they come from distinct disciplines. We recruit participants from a local university through an internal web community with snowball sampling [22]. We interview 12 participants (six visual analytics researchers and six domain researchers; Table 3).

Interview protocol. We interview participants in a semi-structured manner. We first ask participants for consent. In doing so, we clarify that we aim to identify potential risks in participants' reasoning and do not intend to blame participants or other researchers. We then ask participants to describe (1) their expertise in DR, *t*-SNE, and UMAP, (2) their experience and justifications in using *t*-SNE and UMAP, and (3) the difficulties that occur while using these techniques (our questionnaire is in Appendix B). The interviews are conducted via a recorded Zoom call. We compensate participants with approximately 15 USD. All interviews are completed within 40 minutes.

5.2 Analysis Procedure

One coder initially analyzes the interview transcripts using thematic analysis [37]. Each code focuses on describing practitioners' behaviors when analyzing high-dimensional data using DR. The coder extracts participant quotes as initial codes and then hierarchically clusters them into two levels based on semantic similarity. Each lower-level cluster is grouped into a subtheme, which is subsequently abstracted into the final set of themes representing higher-level conceptual categories. Note that we group codes as subthemes only when supported by quotations from

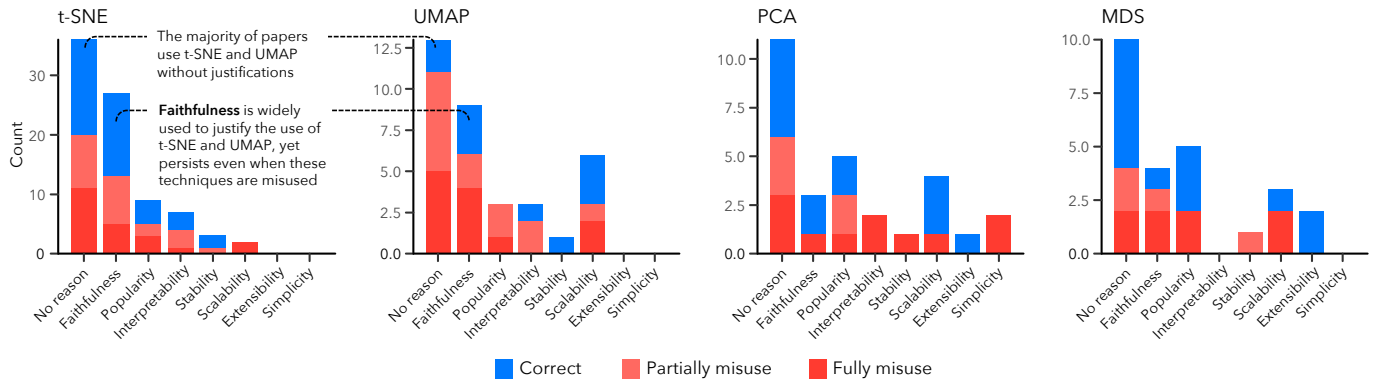

Fig. 6: *The number of appropriate uses and misuses of DR techniques by rationales.* The rationales (x-axis) are sorted in descending order based on the number they are referenced to justify the use of *t*-SNE (leftmost chart).

at least two different participants. Finally, two additional authors and the initial coder collaboratively review and refine the analysis.

5.3 Findings

We identify three themes and elaborate on each as a separate finding.

(Finding 1) Practitioners have limited literacy on DR. We find that the participants have difficulty understanding not only *t*-SNE and UMAP but also other DR techniques, consistent with our takeaways from the literature review (Sect. 4.5). For instance, five participants report difficulties in choosing the final projection for deployment, as the outcomes vary significantly depending on hyperparameter configurations. Three of these participants admit that they are unsure of the proper way to set these hyperparameters. Second, participants often do not have sufficient knowledge of alternative DR techniques. For example, four participants say they are unaware of methods other than *t*-SNE, UMAP, and PCA, and two of these participants mention that they can hardly tell the differences among these three techniques.

We also notice that practitioners' limited literacy makes them "immune" to using *t*-SNE and UMAP for their research. Participants mention that the popularity of *t*-SNE and UMAP makes them less likely to invite criticism of their analytic results or systems. For example, three participants mention using UMAP because they feel that alternative techniques may expose their paper to reviewer criticism. This finding aligns with our observation that many of the papers do not provide adequate justification for using *t*-SNE and UMAP (Sect. 4.4).

(Finding 2) Practitioners receive misleading suggestions. We find that practitioners regularly receive potentially misleading suggestions to use *t*-SNE and UMAP. We identify three primary sources:

Fellow researchers. Five participants indicate that their fellow researchers recommend using *t*-SNE and UMAP. Two participants, in particular, mention that they unreservedly follow recommendations from their advisors or seniors. For example, a participant comments: *"My advisor recommended UMAP, and I used it without verification"*.

Research papers. Two domain researchers state that they used *t*-SNE and UMAP after they frequently encountered them in research papers in their domains (bioinformatics and chemistry). One participant mentions that he regularly uses UMAP because it is frequently cited in recent publications in his domain (bioinformatics) for the same purpose.

Language models. Two participants mention that they ask large language models (LLMs) to recommend a DR technique to use, where the models suggest *t*-SNE, UMAP, and PCA. Both participants report using ChatGPT for this purpose.

These suggestions can be misleading as they typically lack grounded evidence. For example, research papers often misuse *t*-SNE and UMAP (Sect. 4), so relying on these papers can lead to erroneous applications of these techniques. As LLMs are trained on massive text corpora, their frequent recommendations of *t*-SNE and UMAP can be interpreted as reinforcing practitioners' bias toward these techniques. Despite the existence of materials that inform the proper use of DR (Sect. 2.2), practitioners' reliance on such misleading suggestions indicates insufficient motivation to engage with these resources.

Table 4: *The demographics of the participants in our interview study with DR experts.* Exp. and Pub. denote the years of research experience and the number of publications related to DR, respectively.

	Occupation	Age	Gender	Exp.	Pub.	Subdiscipline in DR
P1	Assoc. Prof.	34	Male	12	>10	Faithfulness, Scalability
P2	Assoc. Prof.	48	Male	20	>10	Algorithm, Evaluation
P3	Assist. Prof.	39	Male	13	>10	Interpretation
P4	Assoc. Prof.	40	Male	20	>10	Interpretation
P5	Assist. Prof.	36	Male	11	4	Practical use of DR
P6	Ph.D. Student	26	Male	4	3	Stability, Faithfulness
P7	Ph.D. Student	26	Female	3	3	Faithfulness
P8	Ph.D. Student	28	Male	6	8	Visual analytics for DR

(Finding 3) Practitioners often cherry-pick hyperparameters. We observe another misuse pattern: cherry-picking of hyperparameters. Eight participants report having experience manually tuning the hyperparameters of *t*-SNE and UMAP. They report that they aim to achieve either an interpretable or an aesthetically pleasing projection, with well-separated classes or clusters. Four participants report tuning hyperparameters without understanding their effects on projection results.

5.4 Takeaways

We identify two key takeaways from the interview study:

Practitioners lack understanding of how to use DR properly. The interview study shows that practitioners may have a limited understanding of DR. Many participants (1) hold erroneous beliefs about the faithfulness of *t*-SNE and UMAP (Finding 1), (2) do not know how to select appropriate techniques (Finding 1), and (3) do not know how to properly set hyperparameters (Finding 3).

Practitioners lack motivation to use DR properly. The interview study suggests that practitioners are unaware of how to use DR effectively and lack motivation to do so, aligning with our third takeaway from the literature review (Sect. 4.5). For example, the perception of *t*-SNE and UMAP as "immune to criticism" arises from reviewers' insufficient interest in the proper use of DR.

6 Interview Study with DR Experts

We want to understand why previous efforts are not effective in addressing the misuse (O3) (Sect. 7.1). To this end, we interview visualization researchers whose primary research focus is DR.

6.1 Study Design

Participants and recruitment. We establish two objectives in recruiting the expert researchers. First, we want our experts to have a sufficient **expertise** in (1) the underlying mechanism of DR and (2) how it is used

in practice for visual analytics. Experts without the former may struggle to understand essential concepts for our problem, e.g., the rationale behind different DR techniques and their alignment with analytic tasks. Conversely, experts without the latter may provide limited insights into addressing the misuse. Second, we aim for our pool of experts to be sufficiently **diverse** in their experience with DR.

To achieve these goals, we recruit experts by randomly sending emails to the authors of papers that address the misuse of DR, listed in our related work section (Sect. 2.2). We contact one of the corresponding authors and the first author to further diversify expertise and research experience. Table 4 depicts the demographics of our experts.

Interview protocol. We conduct one-on-one interviews with each expert. One experimenter manages all experiments. After the experts sign the consent form, the experimenter verbally describes the purpose of the experiment and how it will proceed. Then, we share misuse patterns we find from the literature review (Sect. 4) and the interview with practitioners (Sect. 5). Subsequently, we ask interviewees to complete the following questionnaires, asking why the previous efforts in the literature to address the misuse have hardly achieved the goal.

- *"Why do you think these misuses persist despite the large body of literature that informs practitioners not to do so?"*
- *"Why do you think practitioners are not motivated to read existing materials?"*

Each expert answers each question verbally. We wrap up the interview by requesting the experts to share their thoughts on how the misuses can be addressed. We do not limit the interview duration to fully elicit the experts' insights, but all interviews end within 40 minutes. We compensate experts with the equivalent of 20 USD for their participation.

6.2 Analysis Procedure

We apply the same thematic analysis procedure used in the previous interview study (Sect. 5.2), resulting in a set of themes and corresponding subthemes. The only methodological difference is that, given the smaller sample size of eight experts, we permit the formation of a subtheme even when it is supported by a single relevant code.

6.3 Findings

All experts agree that misuse persists despite existing efforts and share the following thoughts on underlying causes.

(Finding 1) Cultivating DR literacy is not easy. Experts note that cultivating DR literacy is not easy, even for trained researchers. Two experts note that although survey papers exist to comprehensively inform the proper use of DR, these resources are difficult for novices to understand. One expert notes that understanding survey papers ultimately requires reading individual papers, which is a demanding task. This finding suggests that existing efforts to instruct practitioners—organizing and presenting information scattered in diverse papers [33] (Sect. 2.2)—hardly reduce the inherent difficulty of learning DR.

(Finding 2) Libraries are promoting the misuse of *t*-SNE and UMAP. Five experts note that highly polished and well-maintained libraries that serve *t*-SNE and UMAP intensify the misuse. They say that although practitioners want to test other DR techniques, they can hardly find and execute implementations, coming back to *t*-SNE and UMAP. Of these, three experts note the need for new libraries that serve diverse DR techniques. Two experts also mention that not only DR techniques but also other infrastructures for utilizing DR, such as preprocessing or evaluation metrics, should be provided as libraries.

(Finding 3) Intrinsic bias of human perception is promoting the misuse of *t*-SNE and UMAP. Two experts casually mention the possibility that previous attempts fail because there exists an intrinsic bias towards well-separated clusters. One of them notes that in such a case, it is natural for *t*-SNE and UMAP to be widely used because these techniques exaggerate cluster structure [31,46]. This is because these techniques are designed to place neighboring points close together and non-neighboring points farther apart than they are in the original space (Sect. 4.3). The finding aligns with the recent finding of Doh et al. [15] that there exists a perceptual factor that makes practitioners prefer DR projections with clearly separated classes or clusters.

(Finding 4) Mitigating the misuse is an urgent problem. Three experts note that we should urgently address these misuses because they may compromise the reliability of scientific discoveries based on visual analytics. P5 states: *"False positives (due to the misuse of DR) are undoubtedly occurring somewhere at this very moment."* Two experts mention that the urgency intensifies because misuses can propagate through citation networks; they emphasize that intervention is needed before these misuses become de facto standards. Both experts note that *t*-SNE and UMAP appear to have already approached this status, warning that mitigation will become increasingly difficult over time.

6.4 Takeaways

We should move beyond papers. We identify the limitations of conventional efforts to mitigate the misuse of DR, which are largely based on academic papers (Finding 1). Moreover, experts emphasize the need for technical solutions to mitigate this bias (Finding 2). This suggests that we need to invest efforts that extend beyond traditional academic approaches to make proper use of DR as the norm.

We need advertisement and maintenance of tools. Though experts note the lack of libraries for alternative DR techniques and related infrastructures, many of these tools already exist. For example, a wide range of alternative DR techniques are available through libraries [1,31,72], and libraries have been released for computing evaluation metrics [27,49]. In addition, approaches to support DR optimization and reduce cherry-picking have been developed, with their implementations made publicly available [17,34]. The discrepancy between the existing tools and our interview findings suggests that these tools remain insufficiently adopted despite their availability. This, in turn, highlights the need for greater dissemination of such tools, as well as maintenance efforts that make them easier to discover, install, and use in practice.

We should act immediately. The interview suggests that immediate action is necessary to address the misuse. Experts warn that the misuse may already have passed the point of no return (Finding 4).

7 Discussion

We discuss future directions to facilitate the proper use of DR.

7.1 Human-Centered Automation

We claim that human-centered automation [65] can be a promising direction to address the misuse of *t*-SNE and UMAP. Our literature review and interviews reveal two underlying causes of misuse: (1) practitioners often lack the knowledge required to use DR appropriately, and (2) they have limited motivation to acquire that knowledge. Addressing these issues and enabling sustained, proper use of DR can benefit from a human-centered use of automation [65]: automating repetitive, redundant tasks (e.g., DR configuration) to make it easier and less burdensome, while preserving practitioners' agency over final decisions (e.g., the selection of DR technique) so they can cultivate DR literacy through the configuration process. We discuss the validity of this argument in a dialectical manner.

An intuitive solution to prevent the misuse of a method is to automate the parts that practitioners struggle to use correctly. For example, AutoML [19] has been widely studied to reduce the burden of machine learning practitioners by automatically tuning hyperparameters. Likewise, it seems beneficial to automate key configuration choices for DR, such as selecting techniques and tuning hyperparameters, in ways that adapt to the analytic tasks and contexts. One plausible direction is to leverage LLMs, given that practitioners regularly ask them to select DR techniques (Finding 2 of Sect. 5.3). For example, equipping LLMs with a knowledge base on DR can enable more reliable recommendations while preserving the convenience of LLM-based assistance.

However, we should remain attentive to the risks of automation [64,65]. If a system makes end-to-end decisions, practitioners may be less motivated to learn to use DR properly, hindering the cultivation of their DR literacy. For example, if hyperparameter tuning is fully automated, practitioners may miss opportunities to learn the underlying

mechanisms of DR techniques. This problem may be amplified by the ease of use of automated solutions, for example, through simplified APIs, which allow users to obtain immediate benefits by applying techniques as-is and may further reduce their motivation to develop deeper DR literacy. This has two disadvantages. First, the lack of DR literacy will limit the ability to critically scrutinize poorly configured DR projections, potentially undermining the credibility of scientific communication. Second, it may also hinder a comprehensive analysis of underlying data structures. The goal of DR-based visual analytics is not to identify a single optimal projection, but to understand a dataset's distribution by comparing various projections while considering their algorithmic properties, appropriate tasks, and context. From this perspective, limited DR literacy may limit practitioners' ability to conduct a thorough analysis even with fully automated DR.

We still do not know the optimal way to use automation to both reduce the technical burden and preserve their agency. We argue that a key direction to address this gap is to assess practitioners' DR literacy and provide adaptive support accordingly. For example, we envision authoring assistants [40, 64] that dynamically infer a user's DR literacy during configuration and guide the process, e.g., by issuing contextual warnings when potentially improper settings are detected. A central challenge is to reliably estimate practitioners' literacy level. Decomposing DR literacy into measurable components and designing assessment methods will be critical steps toward addressing this challenge.

7.2 Facilitating Discourse

We anticipate a natural follow-up question to our warning against the misuse of *t*-SNE and UMAP: *"Which techniques should we use instead?"* Using PCA or MDS for global tasks and *t*-SNE and UMAP for local tasks can serve as a reasonable heuristic. However, this approach is not universally valid because each technique has its own limitations. For instance, the faithfulness of PCA degrades for non-Gaussian data distributions [23, 38]. MDS suffers from quadratic time complexity.

We thus emphasize the importance of encouraging practitioners to cultivate their DR literacy, i.e., to understand the strengths and limitations of diverse DR techniques and to select them based on the given analytical tasks and contexts. For example, specialized techniques such as Densmap preserve density [54] (Fig. 1) while achieving competitive performance in preserving local structure, being suitable for scenarios in which tasks investigating global and local structure are both important [31]. Moreover, other contextual factors play a role: when responsiveness is important (e.g., in real-time streaming data analysis [63]), progressive DR techniques [36] are preferable.

While efforts and studies to enhance DR literacy have been ongoing, our findings indicate that simply publishing guidance or survey papers is not sufficient (Finding 2 in Sect. 4.4; Finding 1 and 2 in Sect. 6.3). In response, we recommend moving beyond traditional text-based dissemination. Organizing workshops, panels, or tutorials at visualization, HCI, and machine learning conferences can help elevate the discourse and emphasize the importance of using DR techniques responsibly. Recent tutorials in EuroVis 2025 [28] and CVPR 2025[1] resonate with this path. Investigating how these efforts synergize with human-centered automation (Sect. 7.1) will also be an intriguing future direction.

7.3 Investigating Domains Outside of Visual Analytics

In this research, we review relevant literature in the field of visual analytics. By doing so, we reveal that visual analytics researchers often use *t*-SNE and UMAP inappropriately. This finding is surprising, as the visual analytics community is arguably expected to have an in-depth understanding of how to use DR appropriately. It further suggests that misuse may be even more prevalent outside the visual analytics community. Our interview with domain researchers partially investigates these fields (Sect. 5), but they cover only four areas: machine learning, HCI, chemistry, and bioinformatics. Examining the use of DR in domains beyond visualization could help us reach a wider audience and uncover more broadly applicable solutions. Cashman et al. [7] recently present a seminal step in this direction by executing a literature review on four domains outside visualization (Physics, Chemistry, Biology, and Business). However, their investigation is based on a literature review and therefore hardly uncovers the in-depth rationales or challenges underlying their findings, unlike our combined approach of literature review and interview study. Conducting additional interviews with domain researchers to probe the underlying causes of the challenges reported in their study will be a promising avenue for future research.

[1] https://cvpr.thecvf.com/virtual/2025/tutorial/35919

7.4 Extending to Other Machine Learning Techniques

We provide a detailed understanding of the phenomenon in which practitioners frequently misuse *t*-SNE and UMAP, often treating these techniques as one-size-fits-all solutions. We find that even experts in data visualization are not immune to these issues (Sect. 5).

We claim to investigate whether a similar tendency exists in the application of other machine learning and AI techniques. For instance, although we cannot always guarantee the faithfulness of LLMs (e.g., due to hallucinations [77]), they are commonly used in many applications without proper performance evaluation, often justified by their perceived ability to generate human-like responses [24]. Investigating whether LLMs are applied to appropriate tasks and properly parameterized is thus necessary. As our research does for *t*-SNE and UMAP, such examinations will help develop solutions that support the proper use of machine learning for visual analytics and stimulate related discourse.

7.5 Immediate Actions

Although we identify future research directions that could help mitigate the misuse of DR, pursuing them will require substantial time and effort. To complement, we provide a simple workflow guideline that can help practitioners immediately avoid misusing DR in visual analytics.

Step 1: Understand your tasks. Practitioners should understand their target task. They can review the analytic tasks we list in Sect. 4.2.2 and Fig. 2. They can then identify the tasks that align with their objectives.

Step 2: Find DR techniques and metrics that match your task. Practitioners are required to find DR techniques and metrics that match their tasks. In terms of techniques, practitioners can naively rely on the alignment between DR techniques and tasks that we provided in Sect. 4.3. Ideally, practitioners can find matching techniques by reviewing papers that propose or benchmark DR techniques, comparing their pros and cons Table 1. For metrics, practitioners can rely on the literature review and taxonomy on evaluation metrics [27, 49].

Step 3: Optimize hyperparameters. In this step, practitioners should iteratively execute the selected DR techniques to optimize their hyperparameters. This is achieved by testing various hyperparameter settings and by evaluating the resulting projections using the selected DR evaluation metric. Recent studies [17, 34] have released code for optimizing hyperparameters efficiently, while traditional approaches like random search and Bayesian optimization [66] can also be used.

Step 4: Execute post-hoc investigation. Even when projections are properly configured and optimized, they cannot fully avoid distortion [29, 56]. Therefore, it is important to understand where and how such distortion occurs [28, 47]. We recommend using distortion visualizations that mark poorly projected points [29, 47]. Such visualizations are already available in existing libraries [27].

8 Conclusion

We critically examine the misuse of *t*-SNE and UMAP in visual analytics. Through the literature review and interview studies, we verify the existence of the misuse, reveal why such misuse occurs. and identify that existing efforts fail to motivate practitioners to develop their DR literacy. Based on these findings, we suggest leveraging human-centered automation to mitigate DR misuse. We also suggest future directions to address the misuse.

Our study highlights the need for broader investigations into the misuse of DR. Although the limited sample size of our studies may raise concerns about the generalizability, such small-scale qualitative studies have already revealed a range of important problems in how DR is used. We thus call for large-scale interview studies and systematic investigations in visual analytics and beyond the field as future work.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2023R1A2C2005209), the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)], and by the SNU-Global Excellence Research Center establishment project. The ICT at Seoul National University provided research facilities for this study. Hyeon is in part supported by the Google Ph.D. Fellowship.

References

[1] E. Amid and M. K. Warmuth. Trimap: Large-scale dimensionality reduction using triplets, 2022. 2, 4, 5, 8

[2] S. Arora, W. Hu, and P. K. Kothari. An analysis of the t-sne algorithm for data visualization. In S. Bubeck, V. Perchet, and P. Rigollet, eds., *31st Conference On Learning Theory*, vol. 75, pp. 1455–1462, 2018. 4

[3] D. Atzberger, T. Cech, W. Scheibel, J. Döllner, M. Behrisch, and T. Schreck. A large-scale sensitivity analysis on latent embeddings and dimensionality reductions for text spatializations. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):305–315, 2025. doi: 10.1109/TVCG.2024.3456308 5

[4] B. C. Benato, A. X. Falcão, and A. C. Telea. Linking data separation, visual separation, classifier performance using multidimensional projections. In *Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 229–255. Springer Nature Switzerland, Cham, 2024. 5

[5] J. Bernard, M. Hutter, M. Zeppelzauer, M. Sedlmair, and T. Munzner. Proseco: Visual analysis of class separation measures and dataset characteristics. *Computers & Graphics*, 96:48–60, 2021. doi: 10.1016/j.cag.2021.03.004 5

[6] M. Brehmer, M. Sedlmair, S. Ingram, and T. Munzner. Visualizing dimensionally-reduced data: Interviews with analysts and a characterization of task sequences. In *Proc. of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, 8 pages, p. 1–8, 2014. doi: 10.1145/2669557.2669559 3

[7] D. Cashman, M. Keller, H. Jeon, B. C. Kwon, and Q. Wang. A critical analysis of the usage of dimensionality reduction in four domains. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–20, 2025. doi: 10.1109/TVCG.2025.3567989 1, 2, 3, 4, 9

[8] T. Chari and L. Pachter. The specious art of single-cell genomics. *PLOS Computational Biology*, 19(8):1–20, 08 2023. doi: 10.1371/journal.pcbi.1011288 1

[9] A. Chatzimparmpas, R. M. Martins, and A. Kerren. t-visne: Interactive assessment and interpretation of t-sne projections. *IEEE Transactions on Visualization and Computer Graphics*, 26(8):2696–2714, 2020. doi: 10.1109/TVCG.2020.2986996 3

[10] A. Chatzimparmpas, F. V. Paulovich, and A. Kerren. Hardvis: Visual analytics to handle instance hardness using undersampling and oversampling techniques. *Computer Graphics Forum*, 42(1):135–154, 2023. doi: 10.1111/cgf.14726 1, 2

[11] L. Chen, C. Cheng, H. Wang, X. Wang, Y. Tian, X. Yue, W. Kam-Kwai, H. Zhang, S. Hong, and Q. Li. Fmlens: Towards better scaffolding the process of fund manager selection in fund investments. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–17, 2024. doi: 10.1109/TVCG.2024.3394745 2

[12] F. Cheng, M. S. Keller, H. Qu, N. Gehlenborg, and Q. Wang. Polyphony: an interactive transfer learning framework for single-cell data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):591–601, 2023. doi: 10.1109/TVCG.2022.3209408 2

[13] A. Coenen and A. Pearce. Understanding umap. *Google PAIR*, 2019. 1, 2, 4, 5, 6

[14] A. D. Col, P. Valdivia, F. Petronetto, F. Dias, C. T. Silva, and L. G. Nonato. Wavelet-based visual analysis of dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 24(8):2456–2469, 2018. doi: 10.1109/TVCG.2017.2746080 3

[15] S. Doh, H. Jeon, S. Shin, G. J. Quadri, N. W. Kim, and J. Seo. Understanding bias in perceiving dimensionality reduction projections, 2025. 4, 8

[16] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, 10 pages, p. 473–482, 2012. doi: 10.1145/2207676.2207741 2

[17] M. Espadoto, R. M. Martins, A. Kerren, N. S. T. Hirata, and A. C. Telea. Toward a quantitative survey of dimension reduction techniques. *IEEE Transactions on Visualization and Computer Graphics*, 27(3):2153–2173, 2021. doi: 10.1109/TVCG.2019.2944182 1, 4, 8, 9

[18] R. Etemadpour, R. Motta, J. G. d. S. Paiva, R. Minghim, M. C. F. de Oliveira, and L. Linsen. Perception-based evaluation of projection methods for multidimensional data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 21(1):81–94, 2015. doi: 10.1109/TVCG.2014.2330617 3

[19] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter. Efficient and robust automated machine learning. *Advances in neural information processing systems*, 28, 2015. 8

[20] T. Fujiwara, Shilpika, N. Sakamoto, J. Nonaka, K. Yamamoto, and K.-L. Ma. A visual analytics framework for reviewing multivariate time-series data with dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, 27(2). doi: 10.1109/TVCG.2020.3028889 3

[21] T. Fujiwara, X. Wei, J. Zhao, and K.-L. Ma. Interactive dimensionality reduction for comparative analysis. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):758–768, 2022. doi: 10.1109/TVCG.2021.3114807 2

[22] L. A. Goodman. Snowball sampling. *The Annals of Mathematical Statistics*, 32(1):148–170, 1961. 6

[23] F. Han and H. Liu. Semiparametric principal component analysis. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds., *Advances in Neural Information Processing Systems*, vol. 25. Curran Associates, Inc., 2012. 9

[24] Y. He, K. Xu, S. Cao, Y. Shi, Q. Chen, and N. Cao. Leveraging foundation models for crafting narrative visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–20, 2025. doi: 10.1109/TVCG.2025.3542504 9

[25] F. Hohman, K. Wongsuphasawat, M. B. Kery, and K. Patel. Understanding and visualizing data iteration in machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 13 pages, p. 1–13, 2020. doi: 10.1145/3313831.3376177 2

[26] H. Hong, S. Yoo, Y. Jin, and Y. Jang. How can we improve data quality for machine learning? a visual analytics system using data and process-driven strategies. In *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*, pp. 112–121, 2023. doi: 10.1109/PacificVis56936.2023.00020 3

[27] H. Jeon, A. Cho, J. Jang, S. Lee, J. Hyun, H.-K. Ko, J. Jo, and J. Seo. Zadu: A python library for evaluating the reliability of dimensionality reduction embeddings. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pp. 196–200, 2023. doi: 10.1109/VIS54172.2023.00048 8, 9

[28] H. Jeon, T. Fujiwara, and R. M. Martins. Reliable Visual Analytics with Dimensionality Reduction: Quality Evaluation and Interpretation of Projections. In *EuroVis 2025 - Panels and Tutorials*. The Eurographics Association, 2025. doi: 10.2312/evt.20251118 9

[29] H. Jeon, H.-K. Ko, J. Jo, Y. Kim, and J. Seo. Measuring and explaining the inter-cluster reliability of multidimensional projections. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):551–561, 2021. doi: 10.1109/TVCG.2021.3114833 9

[30] H. Jeon, H.-K. Ko, S. Lee, J. Jo, and J. Seo. Uniform manifold approximation with two-phase optimization. In *2022 IEEE Visualization and Visual Analytics (VIS)*, pp. 80–84, 2022. doi: 10.1109/VIS54862.2022.00025 1, 2, 3, 4, 5

[31] H. Jeon, K. Ko, S. Lee, J. Hyun, T. Yang, G. Go, J. Jo, and J. Seo. Umato: Bridging local and global structures for reliable visual analytics with dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–18, 2025. doi: 10.1109/TVCG.2025.3602735 8, 9

[32] H. Jeon, Y.-H. Kuo, M. Aupetit, K.-L. Ma, and J. Seo. Classes are not clusters: Improving label-based evaluation of dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):781–791, 2024. doi: 10.1109/TVCG.2023.3327187 1, 2, 3, 4, 5

[33] H. Jeon, H. Lee, Y.-H. Kuo, T. Yang, D. Archambault, S. Ko, T. Fujiwara, K.-L. Ma, and J. Seo. Unveiling high-dimensional backstage: A survey for reliable visual analytics with dimensionality reduction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025. doi: 10.1145/3706598.3713551 1, 2, 3, 8

[34] H. Jeon, J. Park, S. Lee, D. H. Kim, S. Shin, and J. Seo. Dataset-adaptive dimensionality reduction, 2025. 8, 9

[35] H. Jeon, G. J. Quadri, H. Lee, P. Rosen, D. A. Szafir, and J. Seo. Clams: A cluster ambiguity measure for estimating perceptual variability in visual clustering. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):770–780, 2024. doi: 10.1109/TVCG.2023.3327201 4

[36] J. Jo, J. Seo, and J.-D. Fekete. Panene: A progressive algorithm for indexing and querying approximate k-nearest neighbors. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1347–1360, 2020. doi: 10.1109/TVCG.2018.2869149 9

[37] H. Joffe. *Thematic Analysis*, chap. 15, pp. 209–223. John Wiley & Sons, Ltd, 2011. doi: 10.1002/9781119973249.ch15 6

[38] I. Jolliffe. *Principal Component Analysis*. John Wiley & Sons, Ltd, 2005. doi: 10.1002/0470013192.bsa501 9

[39] M. Kahng, P. Y. Andrews, A. Kalro, and D. H. Chau. Activis: Visual exploration of industry-scale deep neural network models. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):88–97, 2018. doi: 10.1109/TVCG.2017.2744718 2

[40] M. H. Kim, Y. Song, Y. Kim, A. Cho, S. Lee, H. Jeon, and J. Seo. Automated pipeline for detecting and analyzing misleading visual elements. In *2025 IEEE 18th Pacific Visualization Conference (PacificVis)*, pp. 346–351, 2025. doi: 10.1109/PacificVis64226.2025.00041 9

[41] D. Kobak and P. Berens. The art of using t-sne for single-cell transcriptomics. *Nature communications*, 10(1):5416, 2019. doi: 10.1038/s41467-019-13056-x 2

[42] D. Kobak and G. C. Linderman. Initialization is critical for preserving global data structure in both t-sne and umap. *Nature Biotechnology*, 39(2):156–157, Feb 2021. doi: 10.1038/s41587-020-00809-z 2

[43] Y.-H. Kuo, T. Fujiwara, C. C.-K. Chou, C.-h. Chen, and K.-L. Ma. A machine-learning-aided visual analysis workflow for investigating air pollution data. In *2022 IEEE 15th Pacific Visualization Symposium (PacificVis)*, pp. 91–100, 2022. doi: 10.1109/PacificVis53943.2022.00018 2

[44] B. C. Kwon, H. Kim, E. Wall, J. Choo, H. Park, and A. Endert. Axisketcher: Interactive nonlinear axis mapping of visualizations through user drawings. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):221–230, 2017. doi: 10.1109/TVCG.2016.2598446 3

[45] J. Lause, P. Berens, and D. Kobak. The art of seeing the elephant in the room: 2d embeddings of single-cell data do make sense. *bioRxiv*, 2024. doi: 10.1101/2024.03.26.586728 2

[46] J. A. Lee and M. Verleysen. Shift-invariant similarities circumvent distance concentration in stochastic neighbor embedding and variants. *Procedia Computer Science*, 4:538–547, 2011. doi: 10.1016/j.procs.2011.04.056 2, 4, 8

[47] S. Lespinats and M. Aupetit. Checkviz: Sanity check and topological clues for linear and non-linear mappings. *Computer Graphics Forum*, 30(1):113–125, 2011. doi: 10.1111/j.1467-8659.2010.01835.x 9

[48] Y. Li, J. Wang, X. Dai, L. Wang, C.-C. M. Yeh, Y. Zheng, W. Zhang, and K.-L. Ma. How does attention work in vision transformers? a visual analytics attempt. *IEEE Transactions on Visualization and Computer Graphics*, 29(6):2888–2900, 2023. doi: 10.1109/TVCG.2023.3261935 1, 2, 3

[49] A. Machado, M. Behrisch, and A. Telea. Extensible TensorFlow Implementations of Projection Quality Metrics. In *VisGap - The Gap between Visualization Research and Visualization Software*. The Eurographics Association, 2025. doi: 10.2312/visgap.20251160 8, 9

[50] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020. doi: 10.48550/arXiv.1802.03426 2, 4

[51] L. Meng, S. van den Elzen, N. Pezzotti, and A. Vilanova. Class-constrained t-sne: Combining data features and class probabilities. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):164–174, 2024. doi: 10.1109/TVCG.2023.3326600 3

[52] M. Moor, M. Horn, B. Rieck, and K. Borgwardt. Topological autoencoders. In *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, pp. 7045–7054, 13–18 Jul 2020. 3, 4, 5

[53] C. Morariu, A. Bibal, R. Cutura, B. Frénay, and M. Sedlmair. Predicting user preferences of dimensionality reduction embedding quality. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):745–755, 2023. doi: 10.1109/TVCG.2022.3209449 4

[54] A. Narayan, B. Berger, and H. Cho. Assessing single-cell transcriptomic variability through density-preserving data visualization. *Nature biotechnology*, 39(6):765–774, 2021. doi: 10.1038/s41587-020-00801-7 1, 2, 3, 4, 5, 9

[55] C. Nolet, A. Lal, R. Ilango, T. Dyer, R. Movva, J. Zedlewski, and J. Israeli. Accelerating single-cell genomic analysis with gpus. *bioRxiv*, 2022. doi: 10.1101/2022.05.26.493607 4

[56] L. G. Nonato and M. Aupetit. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. *IEEE Transactions on Visualization and Computer Graphics*, 25(8):2650–2673, 2019. doi: 10.1109/TVCG.2018.2846735 1, 2, 3, 4, 5, 9

[57] K. F. Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720 2

[58] N. Pezzotti, J. Thijssen, A. Mordvintsev, T. Höllt, B. Van Lew, B. P. Lelieveldt, E. Eisemann, and A. Vilanova. Gpgpu linear complexity t-sne optimization. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1172–1181, 2020. doi: 10.1109/TVCG.2019.2934307 4

[59] I. Ploshchik, A. Chatzimparmpas, and A. Kerren. Metastackvis: Visually-assisted performance evaluation of metamodels. In *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*, pp. 207–211, 2023. doi: 10.1109/PacificVis56936.2023.00030 3

[60] T. Sainburg, L. McInnes, and T. Q. Gentner. Parametric UMAP Embeddings for Representation and Semisupervised Learning. *Neural Computation*, 33(11):2881–2907, 10 2021. doi: 10.1162/neco_a_01434 4

[61] M. Sedlmair, M. Brehmer, S. Ingram, and T. Munzner. Dimensionality reduction in the wild: Gaps and guidance. *Dept. Comput. Sci., Univ. British Columbia*, pp. 1–10, 2012. 3

[62] L. Shi, J. Hu, Z. Tan, J. Tao, J. Ding, Y. Jin, Y. Wu, and P. M. Thompson. Mvnet: Multi-variate multi-view brain network comparison over uncertain data. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4640–4657, 2022. doi: 10.1109/TVCG.2021.3098123 3

[63] Shilpika, T. Fujiwara, N. Sakamoto, J. Nonaka, and K.-L. Ma. A visual analytics approach for hardware system monitoring with streaming functional data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2338–2349, 2022. doi: 10.1109/TVCG.2022.3165348 9

[64] S. Shin, S. Hong, and N. Elmqvist. Visualizationary: Automating design feedback for visualization designers using large language models. *IEEE Transactions on Visualization and Computer Graphics*, 31(10):8796–8813, 2025. doi: 10.1109/TVCG.2025.3579700 8, 9

[65] B. Shneiderman. *Human-centered AI*. Oxford University Press, 2022. 8

[66] J. Snoek, H. Larochelle, and R. P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, vol. 25. Curran Associates, Inc., 2012. 9

[67] C. O. S. Sorzano, J. Vargas, and A. P. Montano. A survey of dimensionality reduction techniques, 2014. 5

[68] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. 2

[69] L. Van Der Maaten, E. O. Postma, H. J. van den Herik, et al. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10(66-71):13, 2009. 5

[70] J. Venna, J. Peltonen, K. Nybo, H. Aidos, and S. Kaski. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *Journal of Machine Learning Research*, 11(13):451–490, 2010. doi: 10.5555/1756006.1756019 4

[71] E. Ventocilla and M. Riveiro. A comparative user study of visualization techniques for cluster analysis of multidimensional data sets. *Information Visualization*, 19(4):318–338. doi: 10.1177/1473871620922166 2, 5

[72] Y. Wang, H. Huang, C. Rudin, and Y. Shaposhnik. Understanding how dimension reduction tools work: An empirical approach to deciphering t-sne, umap, trimap, and pacmap for data visualization. *Journal of Machine Learning Research*, 22(201):1–73, 2021. 2, 5, 8

[73] M. Wattenberg, F. Viégas, and I. Johnson. How to use t-sne effectively. *Distill*, 2016. doi: 10.23915/distill.00002 1, 2, 5, 6

[74] J. Xia, L. Huang, W. Lin, X. Zhao, J. Wu, Y. Chen, Y. Zhao, and W. Chen. Interactive visual cluster analysis by contrastive dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2022. doi: 10.1109/TVCG.2022.3209423 5

[75] J. Xia, Y. Zhang, J. Song, Y. Chen, Y. Wang, and S. Liu. Revisiting dimensionality reduction techniques for visual cluster analysis: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021. doi: 10.1109/TVCG.2021.3114694 2, 3, 4, 5

[76] X. Xuan, X. Zhang, O.-H. Kwon, and K.-L. Ma. Vac-cnn: A visual analytics system for comparative studies of deep convolutional neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2326–2337, 2022. doi: 10.1109/TVCG.2022.3165347 1, 2

[77] Y. Zhang et al. Siren's song in the ai ocean: A survey on hallucination in large language models, 2023. 9

[78] Y. Zhou and T. O. Sharpee. Using Global t-SNE to Preserve Intercluster Data Structure. *Neural Computation*, 34(8):1637–1651, 07 2022. doi: 10.1162/neco_a_01504 2, 4